

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 09/996,620
Source: IFW 16
Date Processed by STIC: 11-16-2004.

ENTERED



IFW16

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/996,620

DATE: 11/16/2004
 TIME: 13:34:57

Input Set : N:\FANTU\US09996620.raw
 Output Set: N:\CRF4\11162004\I996620.raw

SEQUENCE LISTING

1 (1) GENERAL INFORMATION:
 2 (i) APPLICANT: Boodhoo, Amechand
 3 Seehra, Jasbir
 4 Shaw, Gray
 5 Sako, Dianne
 6 (ii) TITLE OF INVENTION: HIGHLY PURIFIED MOCARHAGIN, A COBRA VENOM
 7 PROTEASE, POLYNUCLEOTIDES ENCODING SAME AND RELATED
 8 PROTEASES, AND THERAPEUTIC USES THEREOF
 9 (iii) NUMBER OF SEQUENCES: 22
 10 (iv) CORRESPONDENCE ADDRESS:
 11 (A) ADDRESSEE: Genetics Institute, Inc.
 12 (B) STREET: 87 CambridgePark Drive
 13 (C) CITY: Cambridge
 14 (D) STATE: Massachusetts
 15 (E) COUNTRY: USA
 16 (F) ZIP: 02140
 17 (v) COMPUTER READABLE FORM:
 18 (A) MEDIUM TYPE: Floppy disk
 19 (B) COMPUTER: IBM PC compatible
 20 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 21 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
 22 (vi) CURRENT APPLICATION DATA:
 C--> 23 (A) APPLICATION NUMBER: US/09/996,620
 C--> 24 (B) FILING DATE: 27-Nov-2001
 25 (C) CLASSIFICATION:
 26 (vii) PRIOR APPLICATION DATA:
 27 (A) APPLICATION NUMBER: 09/026,001
 28 (B) FILING DATE: 18-FEB-1998
 29 (viii) ATTORNEY/AGENT INFORMATION:
 30 (A) NAME: Brown, Scott A.
 31 (B) REGISTRATION NUMBER: 32,724
 32 (C) REFERENCE/DOCKET NUMBER: GI5293B
 33 (ix) TELECOMMUNICATION INFORMATION:
 34 (A) TELEPHONE: (617) 498-8224
 35 (B) TELEFAX: (617) 876-5851
 36 (2) INFORMATION FOR SEQ ID NO: 1:
 37 (i) SEQUENCE CHARACTERISTICS:
 38 (A) LENGTH: 30 amino acids
 39 (B) TYPE: amino acid
 40 (C) STRANDEDNESS: single
 41 (D) TOPOLOGY: linear
 42 (ii) MOLECULE TYPE: peptide

RAW SEQUENCE LISTING

DATE: 11/16/2004

PATENT APPLICATION: US/09/996,620

TIME: 13:34:57

Input Set : N:\FANTU\US09996620.raw

Output Set: N:\CRF4\11162004\I996620.raw

```

43 (iii) HYPOTHETICAL: NO
44 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
45 Thr Asn Thr Pro Glu Gln Asp Arg Tyr Leu Gln Ala Lys Lys Tyr Ile
46 1 5 10 15
47 Glu Phe Tyr Val Val Val Asp Asn Val Met Tyr Arg Lys Tyr
48 20 25 30
50 (2) INFORMATION FOR SEQ ID NO: 2:
51 (i) SEQUENCE CHARACTERISTICS:
52 (A) LENGTH: 48 amino acids
53 (B) TYPE: amino acid
54 (C) STRANDEDNESS: single
55 (D) TOPOLOGY: linear
56 (ii) MOLECULE TYPE: peptide
57 (iii) HYPOTHETICAL: NO
58 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
59 Thr Asn Thr Pro Glu Gln Asp Arg Tyr Leu Gln Ala Lys Lys Tyr Ile
60 1 5 10 15
61 Glu Phe Tyr Val Val Val Asp Asn Val Met Tyr Arg Lys Tyr Thr Gly
62 20 25 30
W--> 63 Lys Leu His Val Ile Thr Xaa Xaa Val Tyr Glu Met Asn Ala Leu Asn
64 35 40 45
66 (2) INFORMATION FOR SEQ ID NO: 3:
67 (i) SEQUENCE CHARACTERISTICS:
68 (A) LENGTH: 15 amino acids
69 (B) TYPE: amino acid
70 (C) STRANDEDNESS: single
71 (D) TOPOLOGY: linear
72 (ii) MOLECULE TYPE: peptide
73 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
74 Glu Ala Thr Glu Tyr Glu Tyr Leu Asp Tyr Asp Phe Leu Pro Glu
75 1 5 10 15
77 (2) INFORMATION FOR SEQ ID NO: 4:
78 (i) SEQUENCE CHARACTERISTICS:
79 (A) LENGTH: 15 amino acids
80 (B) TYPE: amino acid
81 (C) STRANDEDNESS: single
82 (D) TOPOLOGY: linear
83 (ii) MOLECULE TYPE: peptide
84 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
85 Gln Ala Thr Glu Tyr Glu Tyr Leu Asp Tyr Asp Phe Leu Pro Glu
86 1 5 10 15
88 (2) INFORMATION FOR SEQ ID NO: 5:
89 (i) SEQUENCE CHARACTERISTICS:
90 (A) LENGTH: 2050 base pairs
91 (B) TYPE: nucleic acid
92 (C) STRANDEDNESS: double
93 (D) TOPOLOGY: linear
94 (ii) MOLECULE TYPE: cDNA
95 (ix) FEATURE:

```

RAW SEQUENCE LISTING

DATE: 11/16/2004

PATENT APPLICATION: US/09/996,620

TIME: 13:34:57

Input Set : N:\FANTU\US09996620.raw

Output Set: N:\CRF4\11162004\I996620.raw

96 (A) NAME/KEY: CDS

97 (B) LOCATION: 78..1940

98 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```

99 AGTCAATAGG AGAAGAGCTC AGGTTGGCTT GGAAGCAGAA AGAGATTCCT GTCCACCACT 60
100 CCAATCCAGG CTCCAAAATG ATCCAAGCTC TCTTGGTAGC TATATGCTTA GCGGTTTTTC 120
101 CATATCAAGG GAGCTCTATA ATCCTGGAAT CCGGGAATGT TAATGATTAT GAAGTAGTGT 180
102 ATCCACAAAA AGTGCCTGCA TTGTCCAAAG GAGGAGTTCA GAATCCTCAG CCAGAGACCA 240
103 AGTATGAAGA TACAATGCAA TATGAATTTT ACGTGAACGG AGAGCCAGTG GTCCTTCACT 300
104 TAGAAAGAAA TAAAGGACTT TTTTCAGAAG ATTACACTGA AACTCATTAT GCCCCTGATG 360
105 GCAGAGAAAT TACAACAAGC TCTCCAGTTC AGGATCACTG CTATTATCAT GGTACATTC 420
106 AGAATGAAGC TGACTCAAGT GCAGTCATCA GTGCATGTGA TGGCTTGAAA GGACATTTCA 480
107 AGCATCAAGG GGAGACATAC TTTATTGAGC CTTTGGAGCT TTCTGACAGT GAAGCCCATG 540
108 CAATATACAA AGATGAAAAT GTAGAAGAAG AGGAAGAGAT CCCCAAAATC TGTGGGGTTA 600
109 CCCAGACTAC TTGGGAATCA GATGAGCCGA TTGAAAAGTC CTCTCAGTTA ACTAATACTC 660
110 CTGAACAAGA CAGGTACTTG CAGGCCAAAA AATACATCGA GTTTTACGTG GTTGTGGACA 720
111 ATGTAATGTA CMGAAAATAC ACCGGCAAGT TACATGTTAT AACAGAAGA GTATATGAAA 780
112 TGGTCAACGC TTAAATACG ATGTACAGAC GTTTGAATTT TCACATAGCA CTGATTGGCC 840
113 TAGAAATTTG GTCCAACGGA AATGAGATTA ATGTGCAATC AGACGTGCAG GCCACTTTGG 900
114 ACTTATTTGG AGAATGGAGA GAAAATAAAT TGCTGCCACG CAAAAGGAAT GATAATGCTC 960
115 AGTTACTCAC GAGCACTGAG TTCAATGGAA CTACTACAGG ACTTGGTTAC ATAGGCTCCC 1020
116 TCTGTAGTCC GAAGAAATCT GTGGCAGTTG TTCAGGATCA TAGCAAAAGC ACAAGCATGG 1080
117 TGGCAATTAC AATGGCCCAT CAGATGGGTC ATAATCTGGG CATGAATGAT GACAGAGCTT 1140
118 CCTGTACTTG TGGTTCTAAC AAATGCATTA TGTCTACAAA ATATTATGAA TCTCTTTCTG 1200
119 AGTTCAGCTC TTGTAGTGTC CAGGAACATC GGGAGTATCT TCTTAGAGAC AGACCACAAT 1260
120 GCATTCTCAA CAAACCCTCG CGCAAAGCTA TTGTTACACC TCCAGTTTGT GGAAATTACT 1320
121 TTGTGGAGCG GGGAGAAGAA TGTGACTGTG GCTCTCCTGA GGATTGTCAA AATACCTGCT 1380
122 GTGATGCTGC AACTTGTAAG CTGCAACATG AGGCACAGTG TGACTCTGGA GAGTGTGTGT 1440
123 AGAAATGCAA ATTTAAGGGA GCAGGAGCAG AATGCCGGGC AGCAAAGAAAT GACTGTGACT 1500
124 TTCCTGAAC CTGCACTGGC CGATCTGCTA AGTGTCCCAA GGACAGCTTC CAGAGGAATG 1560
125 GACATCCATG CCAAAACAAC CAAGGTTACT GCTACAATGG GACATGTCCC ACCTTGACAA 1620
126 ACCAATGTGC TACTCTCTGG GGGCCAGGTG CAAAATATGTC TCCAGGTTTA TGTTTTATGT 1680
127 TGAATGGAA TGCCCGAAGT TGTGGCTTGT GCAGAAAGGA AAATGGCAGA AAGATTCTAT 1740
128 GTGCAGCAAA GGATGTAAAG TGTGGCAGGT TATTTTGTCAA AAAGAAAAAC TCGATGATAT 1800
129 GCCACTGCCC ACTCCATCAA AGGACCCAAA TTATGGAATG GTTGACCTG GAACAAAATG 1860
130 TGGAGTTAAA AAGGTGTGCA GAAACAGGCA ATGTGTAAA GTATAGACAG CCAACTGATC 1920
131 AAGCACTGCT TCTCTCAATT TGATTTTGGG GATCCTCCTT CCAGAAGGCT TTCCTCAAGT 1980
132 CCAAAGAGAC CCATCTGTCT TTATCCTACT AGTAAATCAC TCTTAGCTTT CAAAAAATAA 2040
133 AAAAGTCGAC 2050

```

135 (2) INFORMATION FOR SEQ ID NO: 6:

136 (i) SEQUENCE CHARACTERISTICS:

137 (A) LENGTH: 621 amino acids

138 (B) TYPE: amino acid

139 (C) STRANDEDNESS:

140 (D) TOPOLOGY: linear

141 (ii) MOLECULE TYPE: protein

142 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

143 Met Ile Gln Ala Leu Leu Val Ala Ile Cys Leu Ala Val Phe Pro Tyr

144 1 5 10 15

145 Gln Gly Ser Ser Ile Ile Leu Glu Ser Gly Asn Val Asn Asp Tyr Glu

RAW SEQUENCE LISTING

DATE: 11/16/2004

PATENT APPLICATION: US/09/996,620

TIME: 13:34:57

Input Set : N:\FANTU\US09996620.raw

Output Set: N:\CRF4\11162004\I996620.raw

146		20		25		30								
147	Val	Val	Tyr	Pro	Gln	Lys	Val	Pro	Ala	Leu	Ser	Lys	Gly	Gly
148		35		40		45								
149	Asn	Pro	Gln	Pro	Glu	Thr	Lys	Tyr	Glu	Asp	Thr	Met	Gln	Tyr
150		50		55		60								
151	His	Val	Asn	Gly	Glu	Pro	Val	Val	Leu	His	Leu	Glu	Arg	Asn
152		65		70		75								80
153	Leu	Phe	Ser	Glu	Asp	Tyr	Thr	Glu	Thr	His	Tyr	Ala	Pro	Asp
154				85		90								95
155	Glu	Ile	Thr	Thr	Ser	Ser	Pro	Val	Gln	Asp	His	Cys	Tyr	Tyr
156				100		105								110
157	Tyr	Ile	Gln	Asn	Glu	Ala	Asp	Ser	Ser	Ala	Val	Ile	Ser	Ala
158				115		120								125
159	Gly	Leu	Lys	Gly	His	Phe	Lys	His	Gln	Gly	Glu	Thr	Tyr	Phe
160		130				135								140
161	Pro	Leu	Glu	Leu	Ser	Asp	Ser	Glu	Ala	His	Ala	Ile	Tyr	Lys
162		145				150								160
163	Asn	Val	Glu	Glu	Glu	Glu	Glu	Ile	Pro	Lys	Ile	Cys	Gly	Val
164						165								175
165	Thr	Thr	Trp	Glu	Ser	Asp	Glu	Pro	Ile	Glu	Lys	Ser	Ser	Gln
166				180		185								190
167	Asn	Thr	Pro	Glu	Gln	Asp	Arg	Tyr	Leu	Gln	Ala	Lys	Lys	Tyr
168				195		200								205
169	Phe	Tyr	Val	Val	Val	Asp	Asn	Val	Met	Tyr	Arg	Lys	Tyr	Thr
170				210		215								220
171	Leu	His	Val	Ile	Thr	Arg	Arg	Val	Tyr	Glu	Met	Val	Asn	Ala
172		225				230								240
173	Thr	Met	Tyr	Arg	Arg	Leu	Asn	Phe	His	Ile	Ala	Leu	Ile	Gly
174				245		250								255
175	Ile	Trp	Ser	Asn	Gly	Asn	Glu	Ile	Asn	Val	Gln	Ser	Asp	Val
176				260		265								270
177	Thr	Leu	Asp	Leu	Phe	Gly	Glu	Trp	Arg	Glu	Asn	Lys	Leu	Leu
178				275		280								285
179	Lys	Arg	Asn	Asp	Asn	Ala	Gln	Leu	Leu	Thr	Ser	Thr	Glu	Phe
180				290		295								300
181	Thr	Thr	Thr	Gly	Leu	Gly	Tyr	Ile	Gly	Ser	Leu	Cys	Ser	Pro
182				305		310								320
183	Ser	Val	Ala	Val	Val	Gln	Asp	His	Ser	Lys	Ser	Thr	Ser	Met
184				325		330								335
185	Ile	Thr	Met	Ala	His	Gln	Met	Gly	His	Asn	Leu	Gly	Met	Asn
186				340		345								350
187	Arg	Ala	Ser	Cys	Thr	Cys	Gly	Ser	Asn	Lys	Cys	Ile	Met	Ser
188				355		360								365
189	Tyr	Tyr	Glu	Ser	Leu	Ser	Glu	Phe	Ser	Ser	Cys	Ser	Val	Gln
190				370		375								380
191	Arg	Glu	Tyr	Leu	Leu	Arg	Asp	Arg	Pro	Gln	Cys	Ile	Leu	Asn
192				385		390								400
193	Ser	Arg	Lys	Ala	Ile	Val	Thr	Pro	Pro	Val	Cys	Gly	Asn	Tyr
194				405		410								415

RAW SEQUENCE LISTING

DATE: 11/16/2004

PATENT APPLICATION: US/09/996,620

TIME: 13:34:57

Input Set : N:\FANTU\US09996620.raw

Output Set: N:\CRF4\11162004\I996620.raw

```

195   Glu Arg Gly Glu Glu Cys Asp Cys Gly Ser Pro Glu Asp Cys Gln Asn
196                               420                               425                               430
197   Thr Cys Cys Asp Ala Ala Thr Cys Lys Leu Gln His Glu Ala Gln Cys
198                               435                               440                               445
199   Asp Ser Gly Glu Cys Cys Glu Lys Cys Lys Phe Lys Gly Ala Gly Ala
200                               450                               455                               460
201   Glu Cys Arg Ala Ala Lys Asn Asp Cys Asp Phe Pro Glu Leu Cys Thr
202                               465                               470                               475                               480
203   Gly Arg Ser Ala Lys Cys Pro Lys Asp Ser Phe Gln Arg Asn Gly His
204                               485                               490                               495
205   Pro Cys Gln Asn Asn Gln Gly Tyr Cys Tyr Asn Gly Thr Cys Pro Thr
206                               500                               505                               510
207   Leu Thr Asn Gln Cys Ala Thr Leu Trp Gly Pro Gly Ala Lys Met Ser
208                               515                               520                               525
209   Pro Gly Leu Cys Phe Met Leu Asn Trp Asn Ala Arg Ser Cys Gly Leu
210                               530                               535                               540
211   Cys Arg Lys Glu Asn Gly Arg Lys Ile Leu Cys Ala Ala Lys Asp Val
212                               545                               550                               555                               560
213   Lys Cys Gly Arg Leu Phe Cys Lys Lys Lys Asn Ser Met Ile Cys His
214                               565                               570                               575
215   Cys Pro Leu His Gln Arg Thr Gln Ile Met Glu Trp Leu His Leu Glu
216                               580                               585                               590
217   Gln Asn Val Glu Leu Lys Arg Cys Ala Glu Thr Gly Asn Val Leu Lys
218                               595                               600                               605
219   Tyr Arg Gln Pro Thr Asp Gln Ala Leu Leu Leu Ser Ile
220                               610                               615                               620

```

222 (2) INFORMATION FOR SEQ ID NO: 7:

223 (i) SEQUENCE CHARACTERISTICS:

224 (A) LENGTH: 2297 base pairs

225 (B) TYPE: nucleic acid

226 (C) STRANDEDNESS: double

227 (D) TOPOLOGY: linear

228 (ii) MOLECULE TYPE: cDNA

229 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```

230 GTCGACCAGT CAACAGGAGA AAAGCTCAGG TTGGCTTGGA AGCAGAAAGA GATTCCTGTC      60
231 CACCAGTCCA ATCCAGGCTC CAAAATGATC CAAGCTCTCT TGGTAATTAT ATGCTTAGCG      120
232 GTTTTTCCAT ATCAAGGGAG CTCTATAATC CTGGAATCTG GGAATGTTAA TGATTATGAA      180
233 GTTGTGTATC CACAAAAAGT GCCTGCATTG CTCAAAGGAG GAGTTCAGAA TCCTCAGCCA      240
234 GAGACCAAGT ATGAAGATAC AATGCAATAT GAATTTCAAG TGAATGGAGA GCCAGTAGTC      300
235 CTTCACTTAG AAAGAAATAA AGGACTTTTT TCAGAAGATT AACTTGAAAC TCATTATGCC      360
236 CCTGATGGCA GAGAAATTAC AACAGCCCT CCGGTTTCAGG ATCACTGCTA TTATCATGGT      420
237 TACATTGAGA ATGAAGCTGA CTCAAGTGCA ATCAGTATG CATGTGATGG CTTGAAAGGA      480
238 CATTTCAGAC ATCAAGGGGA GACATACTTT ATTGAGCCCT TGAAGCTTTT CGACAGTGAA      540
239 TCTCATGCAA TCTACAAAGA TGAAAATGTA GAAAACGAGG ATGAGACCCC CGAAACCTGT      600
240 GGGGTAACCG AGACTACTTG GGAGTCAGAT GAGTCCATCG AAAAGACCTC TCAGTTAACT      660
241 AACACTCCTG AACAGACGG GTACTTGCAG GCCAAAAAAT ACATCGAGTT TTACGTGGTT      720
242 GTGGACAACA GAATGTACAG GTATTACAAA CGCAATGAAC CTGCTATAAA AAGAAGAGTA      780
243 TATGAAATGG TCAACGCTGT AAATACGAAG TACAGACCTT TGAAAATTCA CATAACACTG      840
244 ATTGGCCTAG AAATTTGGTC CAACCATGAT AAGTTTGAAG TGAAGCCAGT AGCGGGTGCC      900

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/996,620

DATE: 11/16/2004

TIME: 13:34:58

Input Set : N:\FANTU\US09996620.raw

Output Set: N:\CRF4\11162004\I996620.raw

:23 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
:24 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
:63 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2 after pos.:32